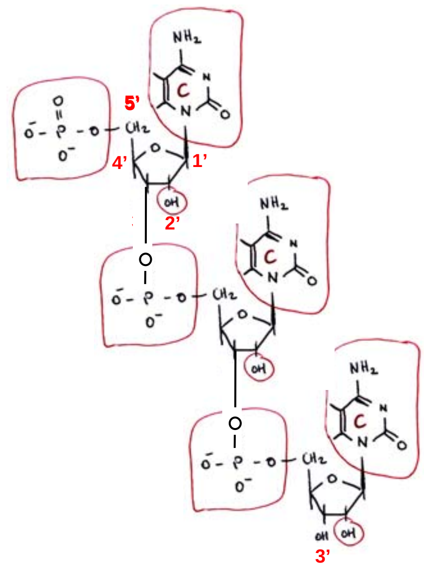
Reading for lecture 2

1. **Structure of DNA and RNA**
2. **Information storage by DNA**
3. **The “Central Dogma”**

- Voet and Voet, Chapters 28 (29,30)
- Alberts et al, Chapters 5 (3)



Structure of DNA and RNA

[Polynucleotides] (2)

5’ and 3’ ends named for numbered carbon atoms in ribose.

De-oxy ribose in DNA, normal ribose in RNA

DNA is more stable as the 2’ OH is a target for reactions that break the chain.

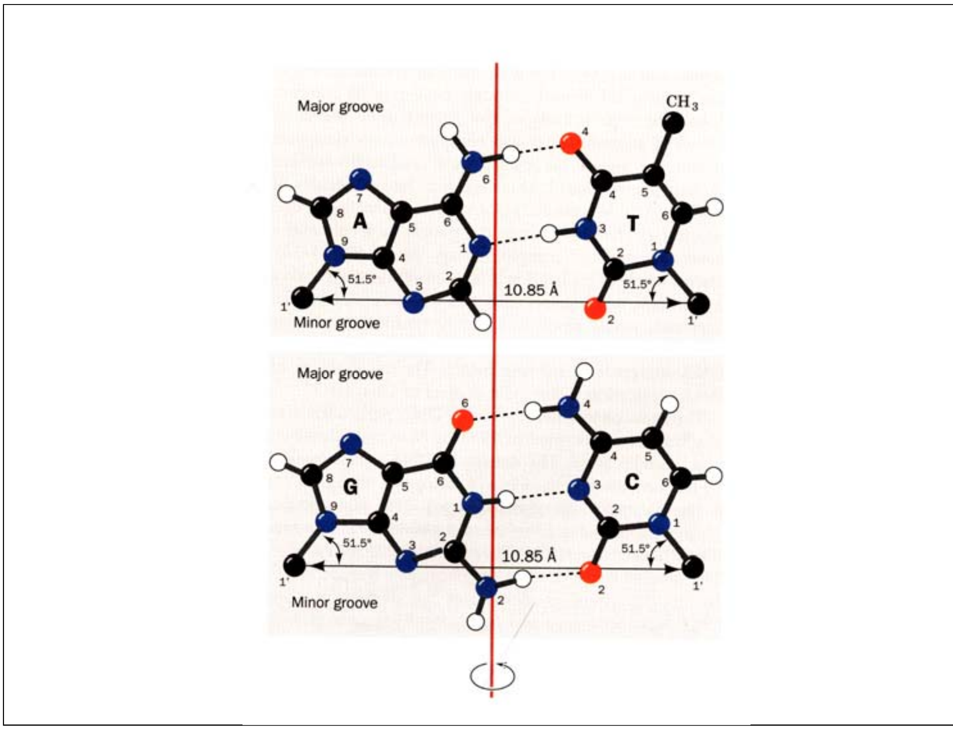
RNA has Uracil (U), DNA has Thymine (T)

3-D Structure of DNA? –early evidence

1940s: Equal amounts of G&C and A&T (G&C range from 25%-75%)

1950s: X-ray diffraction; Helical structure, bases stack with plane normal to long axis.

1953: **Crick and Watson Double Helix** (B-DNA)



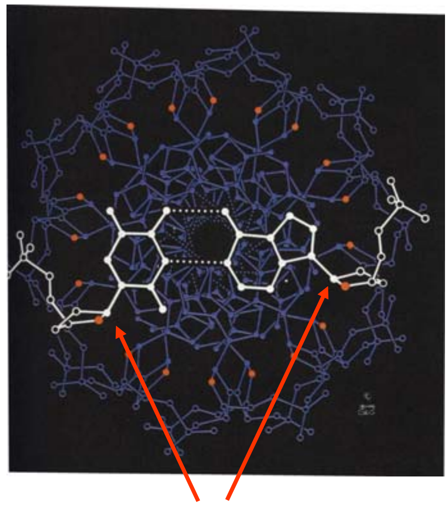
[Base pairs] (3)

AT and GC base pairs by hydrogen bonding. Either pair has same span between 1' carbons.

GC pairs are stronger than AT, 3 vs 2 hydrogen bonds.

Bases planar

Backbone link is slightly longer than base-pair thickness – twist to stack, generates helix



[DNA, end-on] (4)

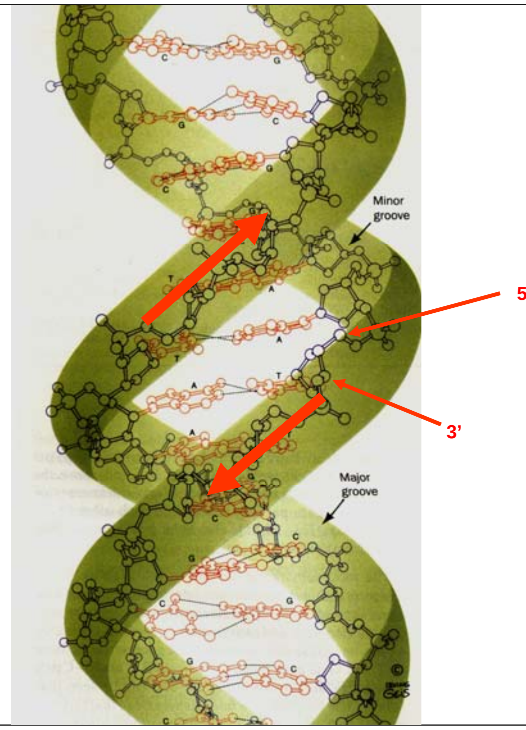
Double helix, strands “glued” to gether by base pairs

Close packed atoms

• length = 3.4 Angstrom per bp

• 10 bp per turn

• Diameter ~ 20 Angstrom



(Windows key + rtclk:

3-D representation of DNA molecule Using “Rasmol”)

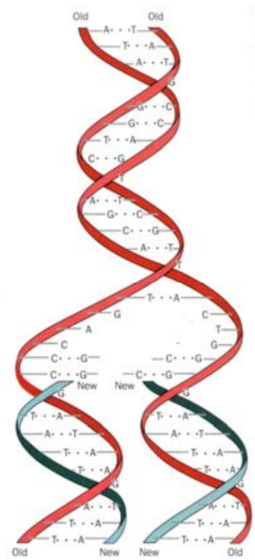
[DNA, side on] (5)

Strands run in opposite directions – 2-fold rotational symmetry

Major and minior grooves because backbones do not span a diameter (see (4)).

[demo: ideal B-DNA in “Rasmol”]

•NB: RNA B-helices are less stable than DNA B-helices because the 2’ OH gets in the way. “steric hindrance”



Information storage by DNA

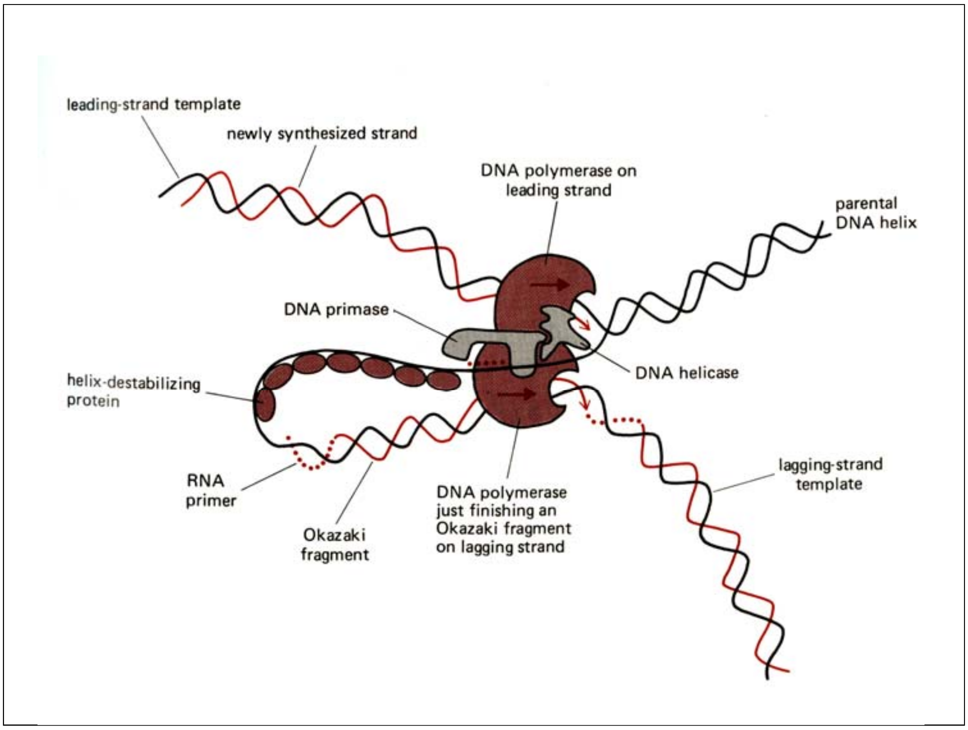
Molecule has a “sequence”. ie) it is a very long word in a 4-letter alphabet.

Sequences are “read” in the 5’ > 3’ direction to reflect the order of polymerisation.

Base-pairing offers a mechanism for **self-directed replication**

[DNA replication] (6)

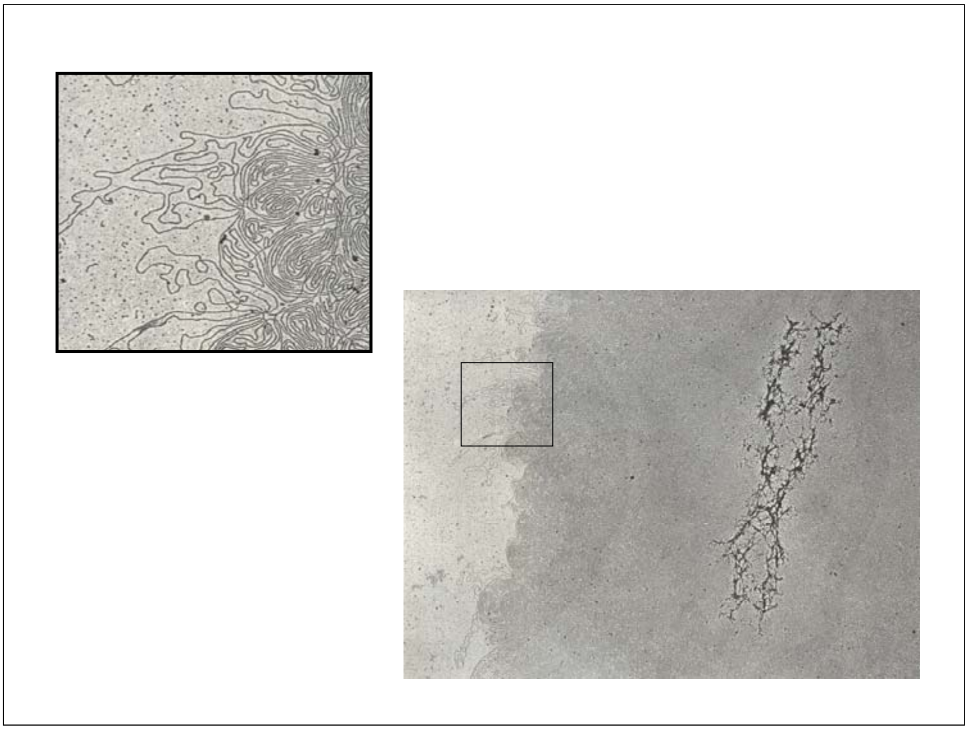
Replication (like most processes) is controlled by molecular machines



[Replication Fork] (7)

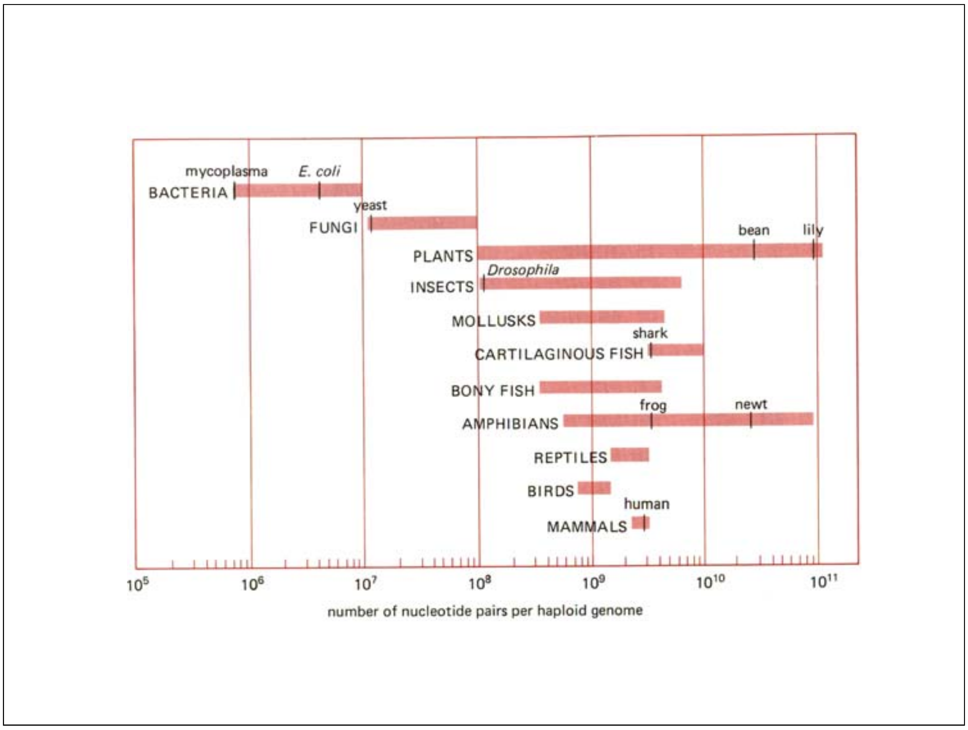
DNA polymerase works 5’ > 3’ on both strands. This introduces asymmetry.

Complicated mechanism. Won’t go into more detail although more is known.



How much DNA is in an organism ?

[Picture of DNA] (8)



[genome sizes] (9)

"Genome"

Diploid vs haploid (two copies allows genetic shuffling to generate variety – "recombination")

Eukaryotes have several linear "Chromosomes" – 23 in humans, each 2 copies.

Prokaryotes usually have one circular DNA molecule

Storage medium	Data element	Element volume	1 GB	
			# elements	volume
CD	Reflect or not 2 options = 1 bit	~ $(5 \times 10^{-7}\ m)^2$ $\times 10^{-3}\ m$ $= 25 \times 10^{-17}\ m^3$	8×10^9	~ $2 \times 10^{-6}\ m^3$ **$(1.25\ cm)^3$**
DNA	Base-pair (bp) 4 options = 2 bits	~ $3.4 \times 10^{-10}\ m$ $\times (20 \times 10^{-10}\ m)^2$ $= 1360 \times 10^{-30}\ m^3$	4×10^9	~ $5.4 \times 10^{-18}\ m^3$ **$(1.7\ \mu m)^3$**
print	Letter 26 options = 4.7 bits	~ $(5 \times 10^{-3}\ m)^2$ $\times 10^{-4}\ m$ $= 2.5 \times 10^{-7}\ m^3$	1.7×10^9	~ $4.25 \times 10^2\ m^3$ **$(7.5\ m)^3$**

[storage capacity] (10)

DNA is extrememly compact – molecular storage

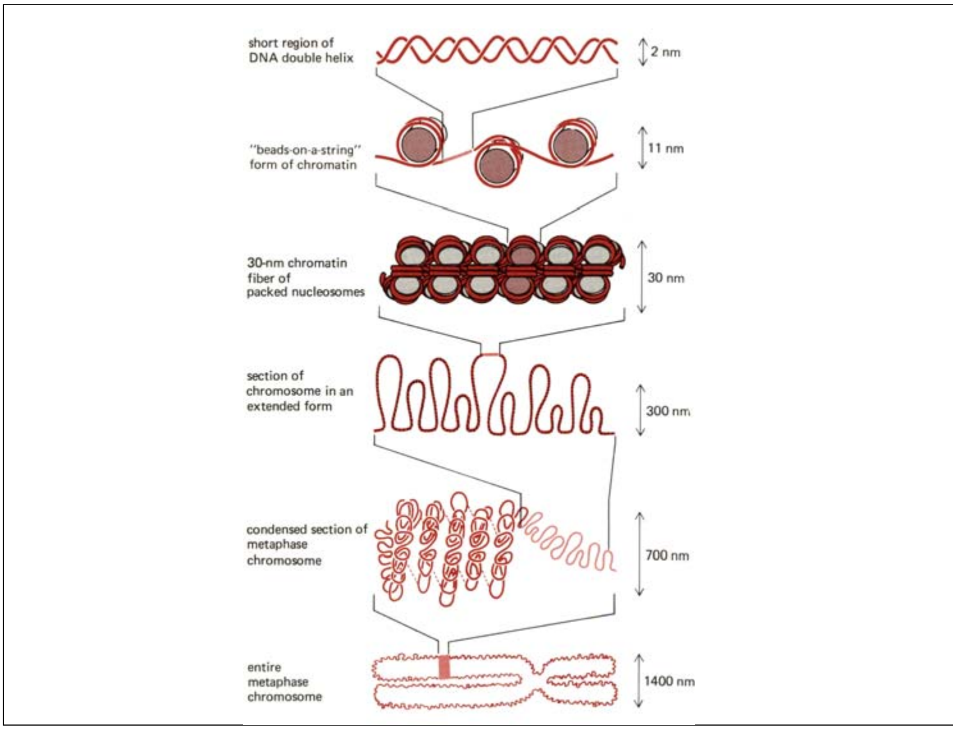
To store the Human Genome : 3.4 billion bp, ~ 1 GB

Volume of DNA:	~ $(1.7\ \mu m)^3$
Actual storage volume (Sperm cell nucleus):	~ $(5\ \mu m)^3$
~1 CD-ROM	~ 10^{11} x greater volume
On paper, 7.5 m cube room	~ 10^{20} x greater volume

[example: physical storage of the human genome] (11)

However, DNA is very fragile. A single linear chromosome is like a 6km long strand of uncooked spaghetti but 0.5 million times smaller. Breaks very easily.

Also problems with knots!



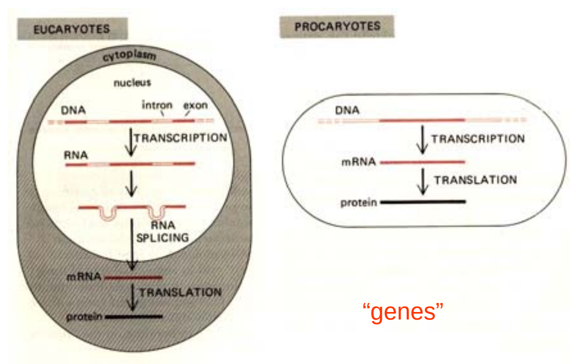
[DNA packaging into chromosomes] (12)

Management of DNA is extensive: Hundreds or thousands of different machines and structural proteins that manipulate it.

The "**Central Dogma**" of molecular biology

DNA directs its own replication and its transcription to RNA which, in turn, directs its translation to proteins.

Francis Crick, 1958



The "Central Dogma"

DNA = blueprint. How is it read /converted into a living, working organism?

[The central dogma] (13)

A **Gene** is a piece of DNA with a particular function

Usually encodes a protein

~30,000-60,000 genes in a human, ~2,000 in *E. coli*

DNA – storage of information

Messenger RNA (mRNA) – "working copy"

Protein – molecular machines and structures

The key fact is that all of these are extended 1-D polymers made of similar subunits.

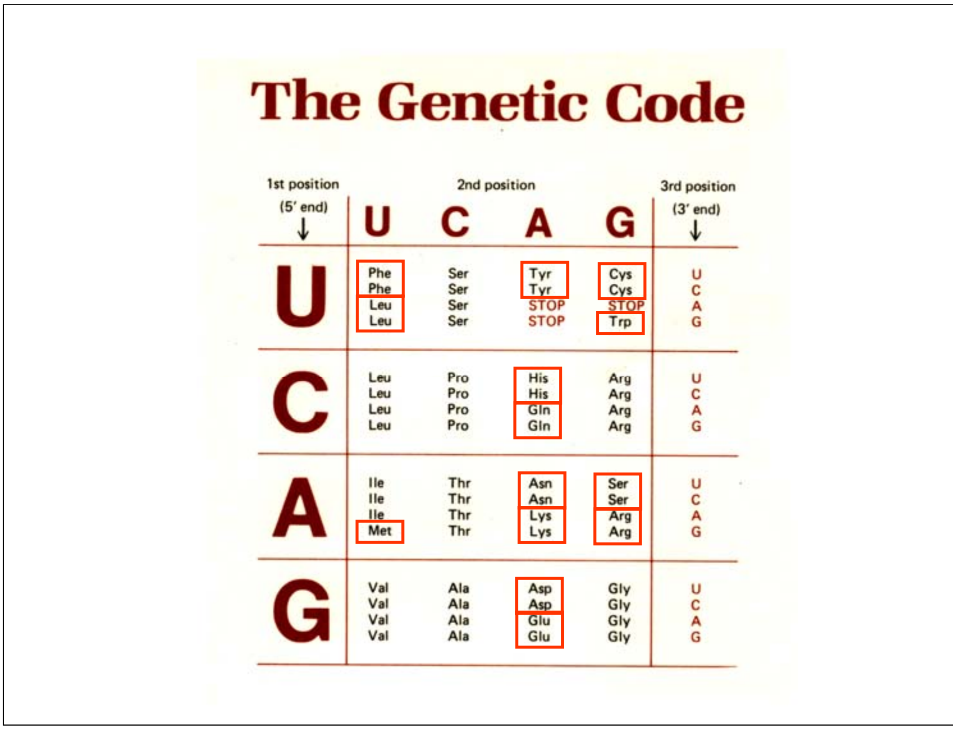
mRNA code is same as DNA except U for T

Proteins are different –

20 amino acids (not 4 bases) so 3-base "codon" is needed (42 = 16 < 20)

Amino Acids and Their Symbols			Codons					
A	Ala	Alanine	GCA	GCC	GCG	GCU		
C	Cys	Cysteine	UGC	UGU				
D	Asp	Aspartic acid	GAC	GAU				
E	Glu	Glutamic acid	GAA	GAG				
F	Phe	Phenylalanine	UUC	UUU				
G	Gly	Glycine	GGA	GGC	GGG	GGU		
H	His	Histidine	CAC	CAU				
I	Ile	Isoleucine	AUA	AUC	AUU			
K	Lys	Lysine	AAA	AAG				
L	Leu	Leucine	UUA	UUG	CUA	CUC	CUG	CUU
M	Met	Methionine	AUG					
N	Asn	Asparagine	AAC	AAU				
P	Pro	Proline	CCA	CCC	CCG	CCU		
Q	Gln	Glutamine	CAA	CAG				
R	Arg	Arginine	AGA	AGG	CGA	CGC	CGG	CGU
S	Ser	Serine	AGC	AGU	UCA	UCC	UCG	UCU
T	Thr	Threonine	ACA	ACC	ACG	ACU		
V	Val	Valine	GUA	GUC	GUG	GUU		
W	Trp	Tryptophan	UGG					
Y	Tyr	Tyrosine	UAC	UAU				

[The genetic code 1] (14)



The Genetic Code

1st position (5′ end)	2nd position: U	C	A	G	3rd position (3′ end)
U	Phe	Ser	Tyr	Cys	U
	Phe	Ser	Tyr	Cys	C
	Leu	Ser	STOP	STOP	A
	Leu	Ser	STOP	Trp	G
C	Leu	Pro	His	Arg	U
	Leu	Pro	His	Arg	C
	Leu	Pro	Gln	Arg	A
	Leu	Pro	Gln	Arg	G
A	Ile	Thr	Asn	Ser	U
	Ile	Thr	Asn	Ser	C
	Ile	Thr	Lys	Arg	A
	Met	Thr	Lys	Arg	G
G	Val	Ala	Asp	Gly	U
	Val	Ala	Asp	Gly	C
	Val	Ala	Glu	Gly	A
	Val	Ala	Glu	Gly	G

[The genetic code 2] (15)

redundancy – 20 amino acids plus stop,. 43 = 64

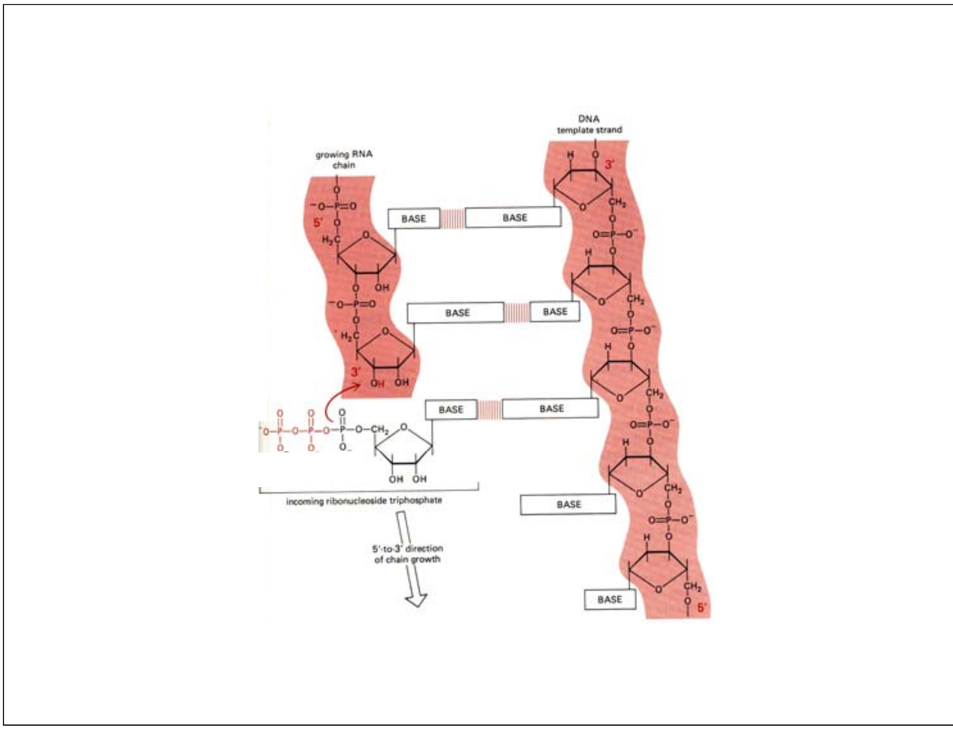
All redundancy is in the third codon

nb: often AG / UC in third codon. NOT to do with base pairing.

NB. P(stop) = 3/64 (bases random) . But coding mRNAs are much longer.

“Open reading frames” – frame is set by the “promoter”

out-of-frame transcripts will be short.



Transcription

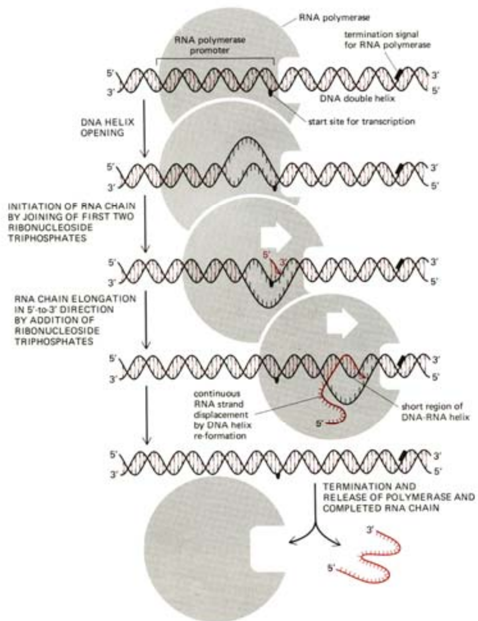
[RNA polymerisation] (16)

Essentially the same as DNA polymerisation

Growing RNA molecule grows from 5’ to 3’

NTP is a high-energy molecule – prevents reverse reaction from happening (both hydrolysis of NTP and subsequently of pyrophosphate are energetically “downhill”.

As always, catalysed by a molecular machine “Enzyme”



[RNA-polymerase] (17)

3 versions in eukaryotes – make different types of RNA

one version in prokaryotes

Large multi-subunit complex. Not fully understood

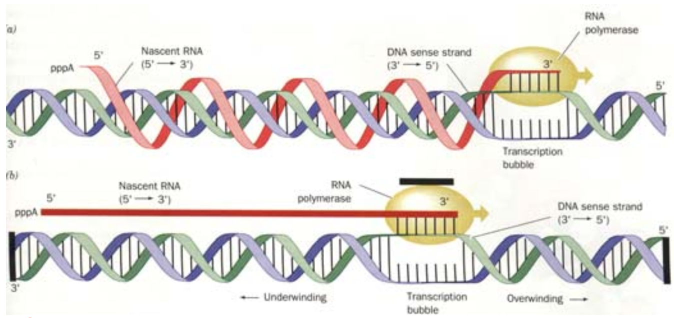
Simple version in “T7 bacteriophage”, single molecule experiments beginning

Promoter: region of DNA with a sequence that binds RNAP more strongly than usual. Also determines which strand/direction to copy.

Elongation: RNAP follows the helix…

Either:

1. RNAP and RNA wind around DNA



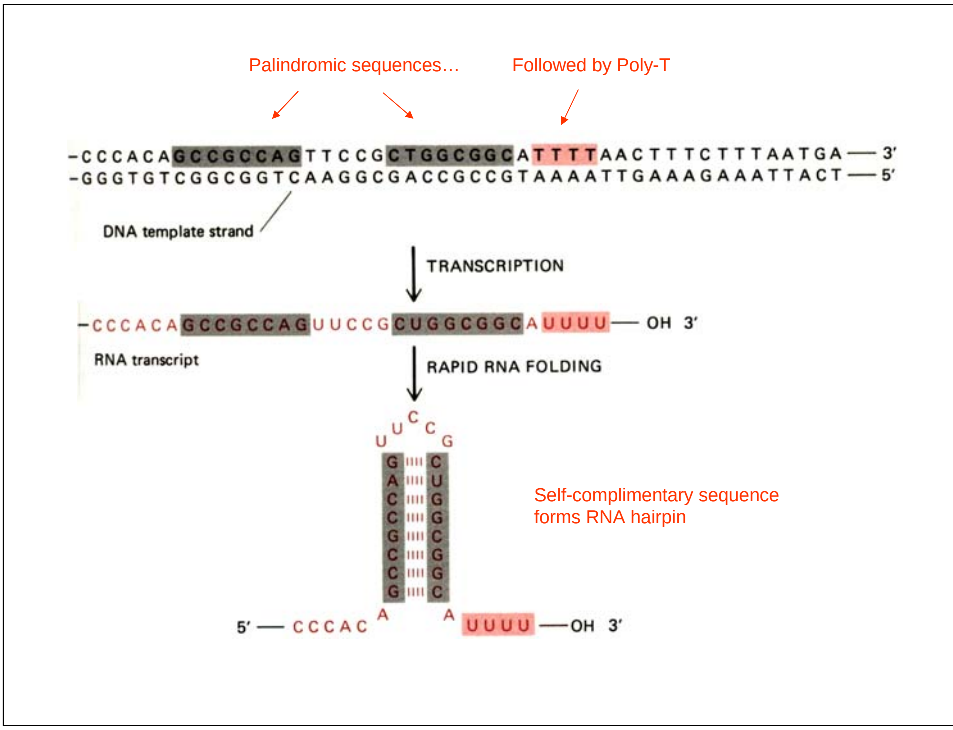
OR:

2. DNA winds / supercoils

[winding or supercoiling] (18)

This will happen in any process that separates the strands of DNA

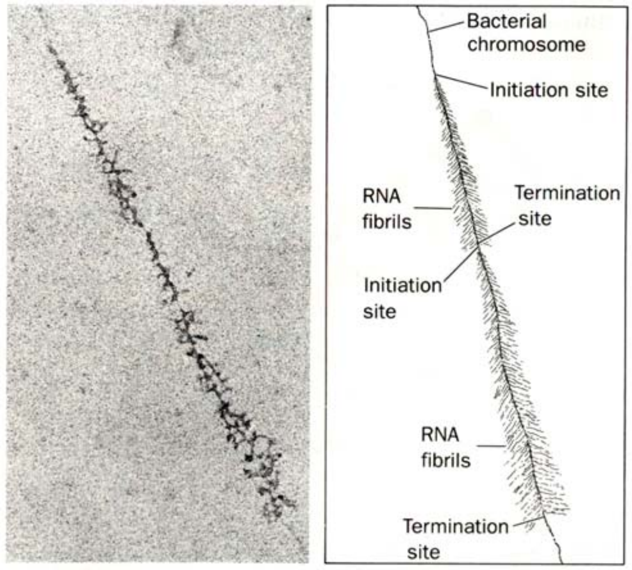
There are families of machines that deal with it by removing supercoils.



[termination] (19)

GC-rich “palindrome” forms strong RNA hairpin – RNA self-hybridisation is important.

Poly A/U region form weak DNA/RNA helix, detaches easily



[picture of transcription] (20)

(next: Translation …)